Elsevier LaTeX template[☆]

Elsevier[1]

Radarweg 29, Amsterdam

Elsevier Inc[a,b], *Global Customer Service*[b,*]

[a]*1600 John F Kennedy Boulevard, Philadelphia*
[b]*360 Park Avenue South, New York*

Abstract

This template helps you to create a properly formatted LaTeX manuscript.

Keywords:

---

1. Main

Suppose (X_1^T, Y_1),...,(X_n^T, Y_n) are i.i.d. from $N_{p+1}(\mu, \Sigma)$, where $X_i \in \mathbb{R}^p$ and $Y_i \in \mathbb{R}$. Denote $X = (X_1, \ldots, X_n)$, $Y = (Y_1, \ldots, Y_n)^T$.

Write $Y = \beta_0 \mathbf{1}_n + X^T\beta + \epsilon$, where $\mathbf{1}_n$ is n dimensional vector with all elements equal to 1. ϵ has distribution $N(0, \sigma^2 I_n)$.

The problem is to test hypotheses $H : \beta = 0$.

Let $Q_n = WW^T$ be the rank decomposition of Q_n, where W_n is a $n \times n-1$ matrix with $W^T W = I_{n-1}$. The new test statistic is

$$T = \frac{y^T Q_n y}{y^T W (W^T X^T X W)^{-1} W^T y}$$

or equivalently

$$\frac{y^T Q_n y}{y^T Q_n (X^T X)^{-1} Q_n y - (y^T Q_n (X^T X)^{-1} \mathbf{1}_n)^2 / (\mathbf{1}_n (X^T X)^{-1} \mathbf{1}_n)}$$

[☆]Fully documented templates are available in the elsarticle package on CTAN.

[*]Corresponding author
Email address: `support@elsevier.com` (Global Customer Service)
URL: `www.elsevier.com` (Elsevier Inc)

[1]Since 1880.

Let $\tilde{y} = W^T y$, $\tilde{X} = XW$, $\tilde{\epsilon} = W^T \epsilon$. Then

$$\tilde{y} = \tilde{X}^T \beta + \tilde{\epsilon}$$

and

$$T = \frac{\tilde{y}^T \tilde{y}}{\tilde{y}^T (\tilde{X}^T \tilde{X})^{-1} \tilde{y}}$$

Next we derive another form of of T. We follow the similar technique of Hotelling's T^2.

Let R be an $(n-1) \times (n-1)$ orthogonal matrix satisfies

$$R\tilde{y} = \begin{pmatrix} \|\tilde{y}\| \\ 0 \\ \cdots \\ 0 \end{pmatrix}.$$

We can write

$$T = \frac{\|\tilde{y}\|^2}{\tilde{y} R^T (R \tilde{X}^T \tilde{X} R^T)^{-1} R \tilde{y}} \tag{1}$$

Denote by $B = R\tilde{X}^T \tilde{X} R^T$, then

$$T = \frac{1}{(B^{-1})_{11}}.$$

Let

$$B = \begin{pmatrix} b_{11} & b_{(1)}^T \\ b_{(1)} & B_{22} \end{pmatrix},$$

and apply the matrix inverse formula, we have $(B^{-1})_{11} = 1/(b_{11} - b_{(1)}^T B_{22}^{-1} b_{(1)})$. Hence

$$T = b_{11} - b_{(1)}^T B_{22}^{-1} b_{(1)}.$$

2. Asymptotic distribution

Note that conditioning on $\tilde{y}$, R is a constant orthogonal matrix. And $\tilde{y}$ is independent of $\tilde{X}$ under null hypotheses. So $B|\tilde{y}$ has the same distribution with

$\tilde{X}^T\tilde{X}$ under null hypotheses. Hence B is independent of $\tilde{y}$ and can be written as

$$B = \sum_{i=1}^{p} \lambda_i z_i z_i^T \tag{2}$$

where z_i's are i.i.d. $n-1$ dimensional random vectors distributed as $N(0, I_{n-1})$, $\lambda_1 \geq \lambda_2 \ldots \geq \lambda_p > 0$ are eigenvalues of Σ_X. Denote by $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$, $Z = (Z_1, \ldots, Z_p)$. Let $Z_{(1)}$ and $Z_{(2)}$ be the first 1 row and last $n-2$ rows of Z, that is

$$Z = \begin{pmatrix} Z_{(1)} \\ Z_{(2)} \end{pmatrix}.$$

Then

$$\begin{aligned} B &= Z\Lambda Z^T \\ &= \begin{pmatrix} Z_{(1)}\Lambda Z_{(1)}^T & Z_{(1)}\Lambda Z_{(2)}^T \\ Z_{(2)}\Lambda Z_{(1)}^T & Z_{(2)}\Lambda Z_{(2)}^T \end{pmatrix}. \end{aligned} \tag{3}$$

Hence

$$\begin{aligned} T &= Z_{(1)}\Lambda Z_{(1)}^T - Z_{(1)}\Lambda Z_{(2)}^T \big(Z_{(2)}\Lambda Z_{(2)}^T\big)^{-1} Z_{(2)}\Lambda Z_{(1)}^T \\ &= Z_{(1)}\big(\Lambda - \Lambda Z_{(2)}^T \big(Z_{(2)}\Lambda Z_{(2)}^T\big)^{-1} Z_{(2)}\Lambda\big) Z_{(1)}^T. \end{aligned} \tag{4}$$

But

$$\begin{aligned} \mathrm{rank}(\Lambda Z_{(2)}^T \big(Z_{(2)}\Lambda Z_{(2)}^T\big)^{-1} Z_{(2)}\Lambda) &= \mathrm{rank}(\Lambda^{\frac{1}{2}} Z_{(2)}^T \big(Z_{(2)}\Lambda Z_{(2)}^T\big)^{-1} Z_{(2)}\Lambda^{\frac{1}{2}}) \\ &= \mathrm{rank}(I_{n-2}) = n-2, \end{aligned} \tag{5}$$

and

$$\begin{aligned} \mathrm{rank}(\Lambda - \Lambda Z_{(2)}^T \big(Z_{(2)}\Lambda Z_{(2)}^T\big)^{-1} Z_{(2)}\Lambda) &= \mathrm{rank}(I_p - \Lambda^{\frac{1}{2}} Z_{(2)}^T \big(Z_{(2)}\Lambda Z_{(2)}^T\big)^{-1} Z_{(2)}\Lambda^{\frac{1}{2}}) \\ &= p-n+2. \end{aligned} \tag{6}$$

Hence

$$T \sim \sum_{i=1}^{p-n+2} \lambda_i(\Lambda - \Lambda Z_{(2)}^T \big(Z_{(2)}\Lambda Z_{(2)}^T\big)^{-1} Z_{(2)}\Lambda)\chi_1^2$$

By Weyl's inequality, we have for $1 \leq i \leq p-n+2$

$$\lambda_i(\Lambda - \Lambda Z_{(2)}^T \big(Z_{(2)}\Lambda Z_{(2)}^T\big)^{-1} Z_{(2)}\Lambda) \leq \lambda_i(\Lambda), \tag{7}$$

and

$$
\begin{aligned}
&\lambda_i(\Lambda - \Lambda Z_{(2)}^T (Z_{(2)} \Lambda Z_{(2)}^T)^{-1} Z_{(2)} \Lambda) \\
\geq &\lambda_{i+n-2}(\Lambda) + \lambda_{p-n+2}(-\Lambda Z_{(2)}^T (Z_{(2)} \Lambda Z_{(2)}^T)^{-1} Z_{(2)} \Lambda) \qquad (8) \\
= &\lambda_{i+n-2}.
\end{aligned}
$$

Hence

$$\sum_{i=n-1}^{p} \lambda_i \chi_1^2 \leq T \leq \sum_{i=1}^{p-n+2} \lambda_i \chi_1^2$$

Note that under condition $\mathrm{tr}\Sigma^4/(\mathrm{tr}\Sigma^2)^2 \to 0$, we have by Liapounoff central limit theorem that

$$\frac{\sum_{i=1}^{p} \lambda_i \chi_1^2 - \mathrm{tr}\Sigma_X}{\sqrt{\mathrm{tr}(\Sigma_X^2)}} \xrightarrow{\mathcal{L}} N(0, 1).$$

And

$$\frac{T - \mathrm{tr}\Sigma_X}{\sqrt{\mathrm{tr}(\Sigma_X^2)}} - \frac{\sum_{i=1}^{p} \lambda_i \chi_1^2 - \mathrm{tr}\Sigma_X}{\sqrt{\mathrm{tr}(\Sigma_X^2)}} = \frac{T - \sum_{i=1}^{p} \lambda_i \chi_1^2}{\sqrt{\mathrm{tr}(\Sigma_X^2)}}, \qquad (9)$$

To prove (9) $\xrightarrow{P} 0$, we only need to prove

$$\mathrm{E}\Big(\frac{\sum_{i=1}^{n-2} \lambda_i \chi_1^2}{\sqrt{\mathrm{tr}(\Sigma_X^2)}}\Big) \to 0,$$

that is

$$\mathrm{E}\Big(\frac{\sum_{i=1}^{n-2} \lambda_i}{\sqrt{\mathrm{tr}(\Sigma_X^2)}}\Big) \to 0. \qquad (10)$$

If λ_i's are bounded below and above, then (10) is equivalent to

$$n/\sqrt{p} \to 0, \qquad (11)$$

or $p/n^2 \to \infty$. We thus obtain the following theorem.

Theorem 1. *Suppose*

$$E\Big(\frac{\sum_{i=1}^{n-2} \lambda_i}{\sqrt{tr(\Sigma_X^2)}}\Big) \to 0,$$

and

$$\frac{tr\Sigma^4}{(tr\Sigma^2)^2} \to 0.$$

Then under null hypotheses, we have

$$\frac{T - tr\Sigma_X}{\sqrt{tr(\Sigma_X^2)}} \xrightarrow{\mathcal{L}} N(0, 1).$$

3. Full Asymptotic Results

$\Sigma_X = P\Lambda P^T$

Non-spike: there's no principal component ($r = 0$). That is, $\lambda_1 = \cdots = \lambda_p$.

Spike: there's r principal components. That is, $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_r \geq \lambda_{r+1} = \cdots = \lambda_p$. Denote by P_1 the first r column of P and P_2 the last $p - r$ column of P.

$$\begin{aligned} Y &= \beta_0 \mathbf{1}_n + X^T \beta + \epsilon \\ &= \beta_0 \mathbf{1}_n + X^T P_1 P_1^T \beta + X^T P_2 P_2^T \beta + \epsilon \end{aligned} \tag{12}$$

In either case, let λ be $\lambda = \lambda_{r+1} = \cdots = \lambda_p$.

PCR try to do regression between Y and (estimated) $X^T P_1$. If P_1 is observed, then the problem is reduced to testing an ordinary regression model. However, it's not the case.

Simply estimating P_1 and invoke classical testing procedure may not be a good idea since the estimation may not be consistent in high dimension. In fact, there may be even no principal component!

In this paper, testing PCR means testing:

H_0: There's no principal component or there's r principal components but $P_1^T \beta = 0$.

H_1; There's r principal components and $P_1^T \beta \neq 0$.

Next we consider:

1. There's no PC.
2. There's r principal components but $P_1 \beta = 0$.

Assumption 1. *X and ϵ are normal distribution.*

$$\begin{aligned} T &= \frac{\beta^T \tilde{X} \tilde{X}^T \beta + 2\beta^T \tilde{X} \tilde{\epsilon} + \tilde{\epsilon}^T \tilde{\epsilon}}{\beta^T \tilde{X} (\tilde{X}^T \tilde{X})^{-1} \tilde{X}^T \beta + 2\beta^T \tilde{X} (\tilde{X}^T \tilde{X})^{-1} \tilde{\epsilon} + \tilde{\epsilon}^T (\tilde{X}^T \tilde{X})^{-1} \tilde{\epsilon}} \\ &= \frac{A_1 + A_2 + A_3}{B_1 + B_2 + B_3} \end{aligned} \tag{13}$$

3.1. circumstance 1

Assumption 2. $r = 0$.

Assumption 3. $n^2/p \to 0$.

3.1.1. Step 1

Independent of data, generate a random p dimensional orthonormal matrix O with Haar invariant distribution. And

$$T = \frac{(O\beta)^T O\tilde{X}{(O\tilde{X})}^T O\beta + 2(O\beta)^T \tilde{X}\tilde{\epsilon} + \tilde{\epsilon}^T\tilde{\epsilon}}{(O\beta)^T O\tilde{X}{((O\tilde{X})}^T O\tilde{X})^{-1}{(O\tilde{X})}^T\beta + 2(O\beta)^T O\tilde{X}{((O\tilde{X})}^T O\tilde{X})^{-1}\tilde{\epsilon} + \tilde{\epsilon}^T{((O\tilde{X})}^T O\tilde{X})^{-1}\tilde{\epsilon}} \tag{14}$$

Note that conditioning on O, $O\tilde{X}$ is a random matrix with each entry independently distributed as $N(0,\lambda)$. Hence O is independent of $O\tilde{X}$. Observe also that $O\beta/\|\beta\|$ is uniformly distributed on the unit ball. We can without loss of generality assuming that $\beta/\|\beta\|$ is uniformly distributed on the unit ball.

3.1.2. Step 2

Independent of data, generate $R > 0$ with R^2 distributed as χ_p^2. Then $\xi = R\beta/\|\beta\|$ distributed as $N_p(0, I_p)$. Note that conditioning on $\tilde{X}$, $\eta = {(\tilde{X}^T\tilde{X})}^{-1/2}\tilde{X}^T\xi$ is distributed as $N_{n-1}(0, I_{n-1})$. Hence η is independent of $\tilde{X}$.

Then

$$\begin{aligned}
T &= \frac{(\|\beta\|/R)^2\xi^T\tilde{X}\tilde{X}^T\xi + 2(\|\beta\|/R)\xi^T\tilde{X}\tilde{\epsilon} + \tilde{\epsilon}^T\tilde{\epsilon}}{(\|\beta\|/R)^2\xi^T\tilde{X}{(\tilde{X}^T\tilde{X})}^{-1}\tilde{X}^T\xi + 2(\|\beta\|/R)\xi^T\tilde{X}{(\tilde{X}^T\tilde{X})}^{-1}\tilde{\epsilon} + \tilde{\epsilon}^T{(\tilde{X}^T\tilde{X})}^{-1}\tilde{\epsilon}} \\
&= \frac{(\|\beta\|/R)^2\eta^T\tilde{X}^T\tilde{X}\eta + 2(\|\beta\|/R)\eta^T{(\tilde{X}^T\tilde{X})}^{1/2}\tilde{\epsilon} + \tilde{\epsilon}^T\tilde{\epsilon}}{(\|\beta\|/R)^2\eta^T\eta + 2(\|\beta\|/R)\eta^T{(\tilde{X}^T\tilde{X})}^{-1/2}\tilde{\epsilon} + \tilde{\epsilon}^T{(\tilde{X}^T\tilde{X})}^{-1}\tilde{\epsilon}} \\
&= \frac{A_1 + A_2 + A_3}{B_1 + B_2 + B_3}
\end{aligned} \tag{15}$$

3.1.3. Step 3: CLT

Similar to the derivation of the distribution of Hotelling's T^2 statistic.

Now we deal with

$$\frac{A_3}{B_3} = \frac{\tilde{\epsilon}^T\tilde{\epsilon}}{\tilde{\epsilon}^T(\tilde{X}^T\tilde{X})^{-1}\tilde{\epsilon}} \tag{16}$$

Let O be an $(n-1)\times(n-1)$ orthogonal matrix satisfies

$$O\tilde{\epsilon} = \begin{pmatrix} \|\tilde{\epsilon}\| \\ 0 \\ \dots \\ 0 \end{pmatrix}.$$

Then

$$\frac{A_3}{B_3} = \frac{(O\tilde{\epsilon})^T O\tilde{\epsilon}}{(O\tilde{\epsilon})^T\left((\tilde{X}O^T)^T\tilde{X}O^T\right)^{-1}O\tilde{\epsilon}} \tag{17}$$

Note that $\tilde{X}O^T$ has the same distribution as $\tilde{X}$ and is independent of O. We have

$$\frac{A_3}{B_3} \sim \frac{1}{((\tilde{X}^T\tilde{X})^{-1})_{11}}. \tag{18}$$

where $((\tilde{X}^T\tilde{X})^{-1})_{11}$ is the first element of $(\tilde{X}^T\tilde{X})^{-1}$. Apply the matrix inverse formula, we have

$$\frac{A_3}{B_3} \sim (\tilde{X}^T\tilde{X})_{11\cdot 2}. \tag{19}$$

Since $\tilde{X}^T\tilde{X} \sim \text{Wishart}_{n-1}(\lambda I_{n-1}, p)$, $(\tilde{X}^T\tilde{X})_{11\cdot 2} \sim \lambda\chi^2_{p-n+2}$. Hence by CLT,

$$\frac{A_3/B_3 - \lambda(p-n+2)}{\lambda\sqrt{2(p-n+2)}} \xrightarrow{\mathcal{L}} N(0,1). \tag{20}$$

But

$$\begin{aligned} \frac{A_3/B_3 - \lambda(p-n+2)}{\lambda\sqrt{2(p-n+2)}} &= \frac{\sqrt{p}}{\sqrt{p-n+2}}\frac{A_3/B_3 - \lambda(p-n+2)}{\lambda\sqrt{2p}} \\ &= \frac{\sqrt{p}}{\sqrt{p-n+2}}\Big(\frac{A_3/B_3 - \lambda p}{\lambda\sqrt{2p}} + \frac{(n-2)}{\sqrt{2p}}\Big). \end{aligned} \tag{21}$$

By Slutsky Theorem, if $n^2/p \to 0$, we have

$$\frac{A_3/B_3 - \lambda p}{\lambda\sqrt{2p}} \xrightarrow{\mathcal{L}} N(0,1) \tag{22}$$

Similar technique can deal with A_1/B_1.

$$\frac{A_1}{B_1} = \frac{\eta^T \tilde{X}^T \tilde{X} \eta}{\eta^T \eta} \sim (\tilde{X}^T \tilde{X})_{11} \sim \lambda \chi_p^2 \tag{23}$$

Hence by CLT,

$$\frac{A_1/B_1 - \lambda p}{\lambda \sqrt{2p}} \xrightarrow{\mathcal{L}} N(0,1). \tag{24}$$

3.1.4. step 4

It's obvious that $A_3 \asymp n$ and $B_1 \asymp \frac{n}{p}\|\beta\|^2$. We already have $A_1/B_1 \asymp p$ and $A_3/B_3 \asymp p$. It follows that $A_1 \asymp n\|\beta\|^2$ and $B_3 \asymp n/p$. And

$$\begin{aligned} A_2 &= O_P(\|\beta\|/\sqrt{p})\eta^T (\tilde{X}^T \tilde{X})^{1/2} \tilde{\epsilon} \\ &= O_P(\|\beta\|/\sqrt{p})\sqrt{\eta^T (\tilde{X}^T \tilde{X})\eta} \\ &= O_P(\|\beta\|/\sqrt{p}) O_P(\sqrt{np}) \\ &= O_P(\sqrt{n}\|\beta\|), \end{aligned} \tag{25}$$

$$\begin{aligned} B_2 &= O_P(\|\beta\|/\sqrt{p})\eta^T (\tilde{X}^T \tilde{X})^{-1/2} \tilde{\epsilon} \\ &= O_P(\|\beta\|/\sqrt{p})\sqrt{\eta^T (\tilde{X}^T \tilde{X})^{-1}\eta} \\ &= O_P(\|\beta\|/\sqrt{p}) O_P(\sqrt{n/p}) \\ &= O_P(\frac{\sqrt{n}}{p}\|\beta\|). \end{aligned} \tag{26}$$

We can deduce that: If $\|\beta\|^2 \to 0$, then

$$T = \frac{\tilde{\epsilon}^T \tilde{\epsilon}}{\tilde{\epsilon}^T (\tilde{X}^T \tilde{X})^{-1} \tilde{\epsilon}} (1 + o_P(1)). \tag{27}$$

Hence $T/(\lambda p) \to 1$. If $\|\beta\|^2 \to \infty$, then

$$T = \frac{\eta^T \tilde{X}^T \tilde{X} \eta}{\eta^T \eta} (1 + o_P(1)). \tag{28}$$

Hence $T/(\lambda p) \to 1$.

3.1.5. Step 5

If $\|\beta\|^2 \to 0$, we have

$$\begin{aligned}
&\Big|\frac{T-\lambda p}{\lambda\sqrt{2p}} - \frac{A_3/B_3-\lambda p}{\lambda\sqrt{2p}}\Big| \\
&= \frac{1}{\lambda\sqrt{2p}}\Big|\frac{A_1+A_2+A_3}{B_1+B_2+B_3} - \frac{A_3}{B_3}\Big| \\
&= \frac{1}{\lambda\sqrt{2p}}\Big|\frac{(A_1+A_2)B_3-(B_1+B_2)A_3}{(B_1+B_2+B_3)B_3}\Big| \\
&= \frac{O_P(1)}{\lambda\sqrt{2p}}\Big|\frac{(O_P(n\|\beta\|^2)+O_P(\sqrt{n}\|\beta\|))O_P(\frac{n}{p})-(O_P(\frac{n}{p}\|\beta\|^2)+O_P(\frac{\sqrt{n}}{p}\|\beta\|))O_P(n)}{n^2/p^2}\Big| \\
&= O_P(\sqrt{p}\|\beta\|^2)+O_P(\frac{\sqrt{p}}{\sqrt{n}}\|\beta\|)
\end{aligned} \tag{29}$$

Hence if $\sqrt{p}\|\beta\|^2 \to 0$ and $\frac{p}{n}\|\beta\|^2 \to 0$, CLT holds.

On the other hand. If $\|\beta\|^2 \to \infty$, we have

$$\begin{aligned}
&\Big|\frac{T-\lambda p}{\lambda\sqrt{2p}} - \frac{A_1/B_1-\lambda p}{\lambda\sqrt{2p}}\Big| \\
&= \frac{1}{\lambda\sqrt{2p}}\Big|\frac{A_1+A_2+A_3}{B_1+B_2+B_3} - \frac{A_1}{B_1}\Big| \\
&= \frac{1}{\lambda\sqrt{2p}}\Big|\frac{(A_2+A_3)B_1-(B_2+B_3)A_1}{(B_1+B_2+B_3)B_1}\Big| \\
&= \frac{O_P(1)}{\lambda\sqrt{2p}}\Big|\frac{(O_P(\sqrt{n}\|\beta\|)+O_P(n))O_P(\frac{n}{p}\|\beta\|^2)-(O_P(\frac{\sqrt{n}}{p}\|\beta\|)+O_P(\frac{n}{p}))O_P(n\|\beta\|^2)}{\frac{n^2}{p^2}\|\beta\|^4}\Big| \\
&= O_P(\frac{\sqrt{p}}{\sqrt{n}}\|\beta\|^{-1})+O_P(\sqrt{p}\|\beta\|^{-2})
\end{aligned} \tag{30}$$

Hence if $\frac{n}{p}\|\beta\|^2 \to \infty$ and $\frac{1}{\sqrt{p}}\|\beta\|^2 \to \infty$, CLT holds.

3.2. circumstance 2

Assumption 4. *$P_1^T\beta = 0$.*

$$
\begin{aligned}
T &= \frac{\beta^T \tilde{X}\tilde{X}^T \beta + 2\beta^T \tilde{X}\tilde{\epsilon} + \tilde{\epsilon}^T \tilde{\epsilon}}{\beta^T \tilde{X}(\tilde{X}^T\tilde{X})^{-1}\tilde{X}^T\beta + 2\beta^T\tilde{X}(\tilde{X}^T\tilde{X})^{-1}\tilde{\epsilon} + \tilde{\epsilon}^T(\tilde{X}^T\tilde{X})^{-1}\tilde{\epsilon}} \\
&= \frac{\beta^T P_2 P_2^T \tilde{X}\tilde{X}^T P_2 P_2^T \beta + 2\beta^T P_2 P_2^T \tilde{X}\tilde{\epsilon} + \tilde{\epsilon}^T \tilde{\epsilon}}{\beta^T P_2 P_2^T \tilde{X}(\tilde{X}^T\tilde{X})^{-1}\tilde{X}^T P_2 P_2^T\beta + 2\beta^T P_2 P_2^T\tilde{X}(\tilde{X}^T\tilde{X})^{-1}\tilde{\epsilon} + \tilde{\epsilon}^T(\tilde{X}^T\tilde{X})^{-1}\tilde{\epsilon}} \\
&= \frac{A_1 + A_2 + A_3}{B_1 + B_2 + B_3}
\end{aligned} \tag{31}
$$

3.2.1. Step 1

Like before, we have $A_3/B_3 \sim (\tilde{X}^T\tilde{X})_{11\cdot 2}$. Denote by $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_p)$. Let $Z = (Z_1, \ldots, Z_p)$ be a $n-1 \times p$ matrix with all elements independently distributed as $N(0,1)$. Let $Z_{(1)}$ and $Z_{(2)}$ be the first 1 row and last $n-2$ rows of Z, that is

$$
Z = \begin{pmatrix} Z_{(1)} \\ Z_{(2)} \end{pmatrix}.
$$

Then

$$
\begin{aligned}
\tilde{X}^T\tilde{X} &\sim Z\Lambda Z^T \\
&= \begin{pmatrix} Z_{(1)}\Lambda Z_{(1)}^T & Z_{(1)}\Lambda Z_{(2)}^T \\ Z_{(2)}\Lambda Z_{(1)}^T & Z_{(2)}\Lambda Z_{(2)}^T \end{pmatrix}.
\end{aligned} \tag{32}
$$

Hence

$$
\begin{aligned}
T &\sim Z_{(1)}\Lambda Z_{(1)}^T - Z_{(1)}\Lambda Z_{(2)}^T (Z_{(2)}\Lambda Z_{(2)}^T)^{-1} Z_{(2)}\Lambda Z_{(1)}^T \\
&= Z_{(1)}\Lambda^{1/2}\big(I_p - \Lambda^{1/2} Z_{(2)}^T (Z_{(2)}\Lambda Z_{(2)}^T)^{-1} Z_{(2)}\Lambda^{1/2}\big)\Lambda^{1/2} Z_{(1)}^T \\
&\leq Z_{(1)}\Lambda^{1/2}\big(I_p - \hat{V}\hat{V}^T\big)\Lambda^{1/2} Z_{(1)}^T.
\end{aligned} \tag{33}
$$

We require $p = o(n^2)$. The principal space is $V = (e_1, \ldots, e_r)$. Then

$$
Z_{(1)}\Lambda^{1/2}\big(VV^T - \hat{V}\hat{V}^T\big)\Lambda^{1/2} Z_{(1)}^T = o(\sqrt{p}) \tag{34}
$$

Note that

$$
Z_{(1)}\Lambda^{1/2}\big(I - VV^T\big)\Lambda^{1/2} Z_{(1)}^T \sim \lambda\chi^2_{p-r} \tag{35}
$$

Hence $T \leq \lambda\chi^2_{p-r} + o(\sqrt{p})$.

On the other hand, the eigenvalues of $\Lambda^{1/2}\big(I_p - \Lambda^{1/2} Z_{(2)}^T (Z_{(2)} \Lambda Z_{(2)}^T)^{-1} Z_{(2)} \Lambda^{1/2}\big)\Lambda^{1/2}$ is no less than $I_p - \Lambda^{1/2} Z_{(2)}^T (Z_{(2)} \Lambda Z_{(2)}^T)^{-1} Z_{(2)} \Lambda^{1/2}$. Hence $T \geq \lambda \chi^2_{p-n+2}$.

Hence $A_3/B_3 \asymp p$ if $p/n \to \infty$.

Another proof:

$$B_3 \leq \tilde{\epsilon}^T (\tilde{X}^T P_2 P_2^T \tilde{X})^{-1} \tilde{\epsilon} \asymp \frac{n}{p-n+2} \tag{36}$$

To get the lower bound, let $P_2^T \tilde{X} = U_2 D_2 V_2^T$ be the SVD of $P_2^T \tilde{X}$, where U_2 is a $(p-r) \times (n-1)$ orthonormal matrix, D_2 is a $(n-1) \times (n-1)$ diagonal matrix and V_2 is a $(n-1) \times (n-1)$ orthonormal matrix. Then

$$\begin{aligned} B_3 &= \tilde{\epsilon}^T (\tilde{X}^T P_1 P_1^T \tilde{X} + \tilde{X}^T P_2 P_2^T \tilde{X})^{-1} \tilde{\epsilon} \\ &= \tilde{\epsilon}^T (\tilde{X}^T P_1 P_1^T \tilde{X} + V_2 D_2^2 V_2^T)^{-1} \tilde{\epsilon} \\ &= \tilde{\epsilon}^T V_2 D_2^{-1} (D_2^{-1} V_2^T \tilde{X}^T P_1 P_1^T \tilde{X} V_2 D_2^{-1} + I_{n-1})^{-1} D_2^{-1} V_2^T \tilde{\epsilon} \\ &\geq \tilde{\epsilon}^T V_2 D_2^{-1} (I_{n-1} - U^* U^{*T}) D_2^{-1} V_2^T \tilde{\epsilon} \\ &= \tilde{\epsilon}^T (\tilde{X}^T P_2 P_2^T \tilde{X})^{-1} \tilde{\epsilon} - \tilde{\epsilon}^T V_2 D_2^{-1} U^* U^{*T} D_2^{-1} V_2^T \tilde{\epsilon} \end{aligned} \tag{37}$$

for some $(n-1) \times r$ random matrix U^*. But

$$\begin{aligned} \tilde{\epsilon}^T V_2 D_2^{-1} U^* U^{*T} D_2^{-1} V_2^T \tilde{\epsilon} &= \mathrm{tr}(U^{*T} D_2^{-1} V_2^T \tilde{\epsilon}\tilde{\epsilon}^T V_2 D_2^{-1} U^*) \\ &\leq \lambda_{\max}(V_2^T \tilde{\epsilon}\tilde{\epsilon}^T V_2) \mathrm{tr}(U^{*T} D_2^{-2} U^*) \\ &\leq \|\tilde{\epsilon}\|^2 r (\lambda_{\min}(D_2))^{-2} \\ &\asymp \frac{n}{p^2} \end{aligned} \tag{38}$$

Hence

$$B_3 = \tilde{\epsilon}^T (\tilde{X}^T P_2 P_2^T \tilde{X})^{-1} \tilde{\epsilon} (1 + O_P(1/p)) \tag{39}$$

Hence

$$\frac{A_3}{B_3} = \lambda \chi^2_{p-r-n+2} (1 + O_P(1/p)) \tag{40}$$

Hence

$$\frac{A_3/B_3 - \lambda(p-r-n+2)}{\lambda\sqrt{2(p-r-n+2)}} \xrightarrow{\mathcal{L}} N(0,1). \tag{41}$$

There's no restriction between n and p!!! Is there anything wrong? If no, it can be generalized to Xu and Zhao's test!!!

3.2.2. Step 2

Note that $P_2^T\tilde{X}$ is an $(p-r)\times(n-1)$ matrix with all elements independently distributed as $N(0,\lambda)$.

$A_1 \asymp n\|P_2^T\beta\|^2$, $A_2 = O_P(\sqrt{n}\|P_2^T\beta\|)$, $A_3 \asymp n$.

$B_3 \asymp n/p$.

As for B_1,

$$\begin{aligned} B_1 &\le \beta^T P_2 P_2^T \tilde{X}(\tilde{X}^T P_2 P_2^T \tilde{X})^{-1} \tilde{X}^T P_2 P_2^T \beta \\ &\asymp \frac{n-1}{p-r}\|P_2^T\beta\|^2 \end{aligned} \tag{42}$$

To get the lower bound, let $P_2^T\tilde{X} = U_2 D_2 V_2^T$ be the SVD of $P_2^T\tilde{X}$, where U_2 is a $(p-r)\times(n-1)$ orthonormal matrix, D_2 is a $(n-1)\times(n-1)$ diagonal matrix and V_2 is a $(n-1)\times(n-1)$ orthonormal matrix. Then

$$\begin{aligned} B_1 &= \beta^T P_2 P_2^T \tilde{X}(\tilde{X}^T P_1 P_1^T \tilde{X} + \tilde{X}^T P_2 P_2^T \tilde{X})^{-1} \tilde{X}^T P_2 P_2^T \beta \\ &= \beta^T P_2 U_2 D_2 V_2^T (\tilde{X}^T P_1 P_1^T \tilde{X} + V_2 D_2^2 V_2^T)^{-1} V_2 D_2 U_2^T P_2^T \beta \\ &= \beta^T P_2 U_2 (D_2^{-1} V_2^T \tilde{X}^T P_1 P_1^T \tilde{X} V_2 D_2^{-1} + I_{n-1})^{-1} U_2^T P_2^T \beta \end{aligned} \tag{43}$$

Note that U_2 is independent of $(V_2, D_2, P_1^T\tilde{X})$, and

$$(D_2^{-1} V_2^T \tilde{X}^T P_1 P_1^T \tilde{X} V_2 D_2^{-1} + I_{n-1})^{-1} \ge I_{n-1} - U^* U^{*T} \tag{44}$$

where U^* is the first r eigenvectors of $D_2^{-1} V_2^T \tilde{X}^T P_1 P_1^T \tilde{X} V_2 D_2^{-1}$ and is independent of U_2. Note also that U_2 is of Haar distribution. Hence

$$\begin{aligned} B_1 &\ge \beta^T P_2 U_2 (I_{n-1} - U^* U^{*T}) U_2^T P_2^T \beta \\ &\asymp \frac{n-1-r}{p-r}\|P_2^T\beta\|^2 \end{aligned} \tag{45}$$

$$\begin{aligned} \text{Upper-Lower} &\le \beta^T P_2 U_2 U^* U^{*T} U_2^T P_2^T \beta \\ &\asymp \frac{r}{p-r}\|P_2^T\beta\|^2 \end{aligned} \tag{46}$$

Hence

$$\begin{aligned} B_1 &= \beta^T P_2 P_2^T \tilde{X}(\tilde{X}^T P_2 P_2^T \tilde{X})^{-1} \tilde{X}^T P_2 P_2^T \beta + O_p(\frac{r}{p-r}\|P_2^T\beta\|^2) \\ &= \beta^T P_2 P_2^T \tilde{X}(\tilde{X}^T P_2 P_2^T \tilde{X})^{-1} \tilde{X}^T P_2 P_2^T \beta (1 + O_P(1/n)) \end{aligned} \tag{47}$$

$$
\begin{aligned}
B_2 &= O_P(1)\sqrt{\beta^T P_2 P_2^T \tilde{X}(\tilde{X}^T\tilde{X})^{-2}\tilde{X}^T P_2 P_2^T \beta} \\
&\leq \lambda_{\min}(\tilde{X}^T\tilde{X})^{-1/2} O_P(1)\sqrt{\beta^T P_2 P_2^T \tilde{X}(\tilde{X}^T\tilde{X})^{-1}\tilde{X}^T P_2 P_2^T \beta}
\end{aligned}
\tag{48}
$$

$$
\lambda_{\min}(\tilde{X}^T\tilde{X}) \geq \lambda_{\min}(\tilde{X}^T P_2 P_2^T \tilde{X}) \asymp p - r \tag{49}
$$

Hence $B_2 = O_P(\frac{\sqrt{n}}{p}\|P_2^T\beta\|)$.

Hence the similar law of large number and CLT holds.

3.2.3. Step 3

$$
\frac{A_1}{B_1} \sim \frac{\chi_p^2}{1 + O_P(1/n)} = \lambda\chi_p^2(1 + O_P(1/n)) \tag{50}
$$

Hence if $\|P_2^T\beta\| \to \infty$ or $\|P_2^T\beta\| \to 0$,

$$
\frac{T}{\lambda p} \xrightarrow{P} 1. \tag{51}
$$

We have

$$
\frac{A_1/B_1 - \lambda p}{\lambda\sqrt{2p}} \sim \frac{\chi_p^2(1 + O_P(1/n)) - p}{\sqrt{2p}} \asymp N(0,1), \tag{52}
$$

if $p = o(n^2)$.

3.3. Consistency of Test

β from normal distribution. Then consistency can be proved. Assume that $\beta \sim N(0, \sigma_\beta^2 I_p)$. Then $\gamma = (\tilde{X}^T\tilde{X})^{-1/2}\tilde{X}^T\beta \sim N(0, \sigma_\beta^2 I_{n-1})$.

$$
\begin{aligned}
T &= \frac{\beta^T\tilde{X}\tilde{X}^T\beta + 2\beta^T\tilde{X}\tilde{\epsilon} + \tilde{\epsilon}^T\tilde{\epsilon}}{\beta^T\tilde{X}(\tilde{X}^T\tilde{X})^{-1}\tilde{X}^T\beta + 2\beta^T\tilde{X}(\tilde{X}^T\tilde{X})^{-1}\tilde{\epsilon} + \tilde{\epsilon}^T(\tilde{X}^T\tilde{X})^{-1}\tilde{\epsilon}} \\
&= \frac{\gamma^T\tilde{X}^T\tilde{X}\gamma + 2\gamma^T(\tilde{X}^T\tilde{X})^{1/2}\tilde{\epsilon} + \tilde{\epsilon}^T\tilde{\epsilon}}{\gamma^T\gamma + 2\gamma^T(\tilde{X}^T\tilde{X})^{-1/2}\tilde{\epsilon} + \tilde{\epsilon}(\tilde{X}^T\tilde{X})^{-1}\tilde{\epsilon}} \\
&= \frac{A_1 + A_2 + A_3}{B_1 + B_2 + B_3}
\end{aligned}
\tag{53}
$$

$A_1 \sim \|\gamma\|^2\sum_{i=1}^p \lambda_i\chi_1^2 \asymp \|\gamma\|^2(p + \lambda_1) \asymp \sigma_\beta^2 n(p + \lambda_1)$. $A_2 = O_P(\sqrt{A_1})$. $A_3 \asymp n$.

$B_1 \asymp \sigma_\beta^2 n$. $B_3 \leq \tilde{\epsilon}(\tilde{X}^T P_2 P_2^T \tilde{X})^{-1}\tilde{\epsilon} \asymp n/p$. $B_2 = O_P(\sqrt{B_3}\sigma_\beta)$.

$A_1/B_1 \sim \sum_{i=1}^p \lambda_i\chi_1^2$.

4. Simulation Results

References